



ИГОРЬ АШМАНОВ

# БОЛЬШИЕ ПОЛЬЗОВАТЕЛЬСКИЕ ДАННЫЕ: РИСКИ И ВОЗМОЖНОСТИ

ПОДГОТОВЛЕНО  
«АШМАНОВ И ПАРТНЕРЫ»  
НОЯБРЬ 17



# БОЛЬШИЕ ПОЛЬЗОВАТЕЛЬСКИЕ ДАННЫЕ – НОВАЯ РЕАЛЬНОСТЬ

- ❖ Накоплены большие объёмы БПД, возникли новейшие технологии обработки – нейронные сети, машинное обучение
- ❖ Развился огромный рынок, в условиях правового вакуума
- ❖ Происходит массовая утечка БПД за рубеж
- ❖ В США в 2017 разрешили свободную продажу БПД операторами
- ❖ В ЕС в 2016 введена General Data Protection Regulation (GDPR)
- ❖ Пользовательский профиль оценивается в 10-20 долларов
- ❖ В РФ крупные компании создают СРО по БПД, готовится закон о БПД

**Есть большие риски и большие возможности**



# ЧТО ТАКОЕ «БОЛЬШИЕ ДАННЫЕ»

- Модное, «шумовое» слово, не научное, а скорее медийное понятие
- Не новое понятие (ему примерно 30 лет, раньше называлось Massive Data Processing, Data Mining и т.п.)
- Не алгоритмы обработки, а сами данные
- Большие массивы данных со специфическими свойствами



# «БОЛЬШИЕ ДАННЫЕ» - ОПРЕДЕЛЕНИЕ

- Большие (нельзя просмотреть и обработать вручную)
- **Целое больше суммы частей** (сверху видны закономерности и связи)
- Данные из разных каналов (соцсети, поисковики, камеры, смартфоны, счётчики)
- Данные разного формата (текст, геоданные, связи, картинки, видео, голос)
- С ретроспективой (накопление истории, признаков и маркеров)
- **Данные о людях**



# ЧТО ТАКОЕ БОЛЬШИЕ ПОЛЬЗОВАТЕЛЬСКИЕ ДАННЫЕ

## Неформальное определение:

**Большие пользовательские данные** – это большие данные, собираемые в сети различными сервисами, устройствами и приложениями о поведении пользователей, понимаемых как пользовательские профили.

- ❖ Бывают первичные (сырые) и производные (вычисляемые)
- ❖ Производные БПД при обработке порождают ПД в том числе



# ПОЧЕМУ БПД – ЭТО НЕ АЛГОРИТМЫ

- Алгоритмы обработки кардинально **меняются** каждые 2-3 года
- В данных важна **ретроспектива**, их никто не удаляет, они накапливаются годами, прямо сейчас
- Данные **перепродаются** через многие звенья рынка и через границы государств, конечные покупатели часто неизвестны
- БПД используют в частных компаниях, которые **не раскрывают способы использования** и алгоритмы обработки.

Через 3-5 лет БПД будут использовать по-другому, другие игроки – а сегодняшние и завтрашние данные сохранят свою ценность



# КАТЕГОРИИ БПД С ТОЧКИ ЗРЕНИЯ ПОЛЬЗОВАТЕЛЯ

- **Явно отдаваемые.** Адреса почты, посты/комментарии, профили, фото, лайки и т.п. «Регулируются» пользовательскими соглашениями.
- **Ограниченного распространения.** Переписка, закрытые обсуждения, география, биометрия.
- **Косвенные данные.** Поведенческие, география, биометрия, фото, видео, привычки, взгляды, граф связей.
- **Воруемые данные.** Фото, видео, СМС, голос, клавиатурный ввод, пароли, переписка (смартфоны, смарт-ТВ, keyloggers, трояны)



# ЧЕМ ОТЛИЧАЮТСЯ ОТ ПД

## Персональные данные

- ❖ **Идентифицирующие.** ФИО, номер кредитки, телефон, адрес, прочее.
- ❖ **Есть осознание проблемы** государством и обществом, есть процедуры хранения и передачи.
- ❖ **Собираются операторами ПД,** с уведомлением пользователя.
- ❖ **Регулируются:** 152-ФЗ, проверки Роскомнадзора.

## БПД

- ❖ **Анонимные/агрегированные/сырые.** Вычисления отделены от сбора, сырые данные от обработанных
- ❖ **Позволяют вычислить ПД:** правило 4 точек в разных сферах, сличение векторов из разных каналов
- ❖ **Собираются и продаются всеми:** реклама, скоринг, коллекторы, спецслужбы, хакеры
- ❖ **Не регулируются в РФ.** Нет осознания проблемы обществом и государством





# КТО СОБИРАЕТ БПД

- **Навигаторы:** поисковики и браузеры (запросы, интересы, клики, посещения третьих сайтов, география)
- **Счётчики** разных типов (до 50 на странице, клики, география, интересы, запросы)
- **Рекламные сети** (интересы, клики, география, покупки)
- **Коммуникационные сервисы:** социальные сети, публичные почты, мессенджеры (соцдем, адреса, фото, интересы, граф связей, видео)
- **Платёжные системы:** финансовая и экономическая информация
- **Платформы:** ОС, провайдеры, платёжные системы (запросы, видео, аудио, тексты, сообщения, аккаунты, пароли)



# НОВЫЕ СУЩНОСТИ

- **Смартфоны и приложения** (тексты, СМС, снимки, геоположение, контакты, ...)
- **Смарт-ТВ:** поток данных из квартиры (видео, жесты, трёхмерные сцены, звук, потребительское поведение).
- **Носимые устройства, фитнес-браслеты** – изначально «в облаке» (биометрия, медицинское состояние, перемещения, паттерны занятий: сон, бег, работа, еда, секс)
- **Анализ ДНК** – данные о национальности, родственниках, болезнях/предрасположенности, расе, этносе
- **Автомобили** (видео, звук, география, поведенческие данные)
- **«Интернет вещей»** - данные об окружении, устройствах, производстве, ремонте, география, звук, видео



# КАКИЕ ДАННЫЕ СОБИРАЮТСЯ

- **Поведенческие:** поисковые запросы, клики, посещения сайтов, география, перемещения, точки пребывания
- **Социальные:** посты, комменты, лайки, соцдем, фото, видео, политические взгляды, работа, привычки, хобби
- **Биологические:** фото, видео, ДНК, отпечатки пальцев, пульс, давление, физическая активность, сигнатура голоса и т.п.
- **Финансово-экономические:** счета, активы, покупки, транзакции, уровень доходов, активы, контрагенты
- **Личные, интимные:** семья, связи, круг общения, привычки, половая ориентация, развлечения, психология, взгляды



# МЕТОДЫ СВЕДЕНИЯ ДАННЫХ И АНАЛИЗА

- **Анализ по всей линейке сервисов:** поиск, новости, рекламная система, почта, социальная сеть, карта, мессенджер, платёжная система, браузер, фотохостинг, счётчик, ...
- **Кросс-канальный анализ:** сравнение векторов с разных источников, правило 4 точек для разных сфер, сведение времён выхода в сеть и постов в соцсетях и т.п.
- **Кросс-платформенный анализ** - сведение воедино разных устройств пользователя (отпечатки браузера, ультразвук и т.п.)



# РИСКИ БПД

- **Слежка:** постоянная утечка БПД граждан РФ к частным компаниям и за рубеж. Большинство платформ - американские
- **Нарушение приватности:** продажа пользовательских данных, навязчивая реклама, раскрытие интимных данных
- **Влияние и пропаганда:** персональное и массовое информационное давление, создание «информационных пузырей»
- **Решение США свободно продавать БПД:** они считают своей юрисдикцией весь мир – значит, будут торговать данными наших граждан в том числе.



# ЭТИЧЕСКИЕ И ПРАВОВЫЕ АСПЕКТЫ

- Проблема информированности пользователей
- Проблема беспечности
- Коммерческий интерес собирателей БПД
- Проблема «наблюдающих сервисов»
- Отсутствие этического кодекса и саморегуляции отрасли
- Отсутствие регулирования оборота, правовой вакуум
- Отсутствие отраслевого регулятора и операторов БПД



# КОМУ ДОЛЖНЫ ПРИНАДЛЕЖАТЬ БПД?

- **Пользователю?** Он не знает, кто собирает, что собирают, не может контролировать или запретить сбор и использование
- **Площадкам?** Они по факту владеют БПД, но их пользовательские соглашения незаконны; их коммерческий интерес – хищнический
- **Вывод: это собственность нации,** нужна защита государством.
- **Метафора «цифровых недр».** Владелец площадки не владеет недрами под ней. Для разработки и оборота «цифровой нефти» нужно регулирование.



# ВОЗМОЖНОСТИ БПД

- **Алгоритмы анализа:** промышленная революция, нельзя упустить, чтобы потом не импортировать
- **Знания о нации:** медицина, образование, настроения, выборы, занятость, развлечения, отдых и т.п.
- **Знания об инфраструктуре и производстве:** транспорт, дороги, ЖКХ, электросети, производства.
- **Знания об экономике:** доходы, структура расходов и сбережений, страхование, льготы, планирование развития, прочее
- **Рынок данных:** профиль пользователя стоит денег, при упорядоченном обороте БПД – очень большой рынок





# ЧТО ДЕЛАТЬ: ГОСУДАРСТВО

- **Законотворчество:** регулирование оборота БПД, аналогия «цифровых недр», категории операторов БПД и их права/обязанности
- **Создание рынка:**
  - стандарты доступа к данным, недискриминирующий доступ к стерилизованным данным
  - равные условия для всех игроков отрасли (регулирование зарубежных игроков)
- **Развитие:** кадры, образование, институты развития, доступ к данным, вычислительная мощь для учёных и бизнеса
- **Работа с пользователями:** просвещение, информирование, создание сервисов управления своими БПД



# ЧТО ДЕЛАТЬ: ИТ-ОТРАСЛЬ

- **Ответственное и правовое поведение:** прекратить попытки ввести своё право на пользовательские данные
- **Саморегулирование:**
  - Стандартные пользовательские соглашения
  - Отраслевая СРО по большим данным
  - Этический кодекс операторов БПД
- **Стандарты информирования** пользователей и государства об обороте БПД
- **Стандарты доступа к данным, стандарты обезличивания**



# СПАСИБО!

# ВОПРОСЫ?

ИГОРЬ АШМАНОВ

Управляющий партнер  
«Ашманов и партнеры»

+7 495 269 0630

+7 985 222 5451

[igor@ashmanov.com](mailto:igor@ashmanov.com)